Ann Finance (2009) 5:189–208
DOI 10.1007/s10436-008-0097-3

RESEARCH ARTICLE

On the calibration of structural credit spread models

Howard Qi · Sheen Liu · Chunchi Wu

Received: 18 April 2007 / Accepted: 26 February 2008 / Published online: 14 March 2008

Abstract Empirical findings are mixed about the performance of structural models for term structure of credit spreads. It is commonly believed that all structural models have equally poor performance after calibration. However, proper calibration is not a trivial issue, especially for highly structural models. This paper proposes a more accurate procedure for calibrating two models: Leland–Toft (J Finance 51:987–1019, 1996) and Collin-Dufresne and Goldstein (J Finance 56:2177–2208, 2001). Using rating-based bond data, we find that the Leland–Toft model has significantly greater explanatory power for credit spreads across rating categories than previously reported. We provide theoretical explanations for these findings, and further extend our empirical analysis to include 286 individual senior bonds. Our findings help clarify the controversies over the performance of structural models in general and that of the Leland–Toft model in particular. In addition, we offer a rigorous procedure that can be used for calibrating other structural models more effectively.

Keywords Calibration · Term structure · Capital structure · Default · Credit spread

JEL Classification G12 · G13 · G30 · G32

H. Qi (✉)
School of Business and Economics, Michigan Technological University,
Houghton, MI 49931, USA
e-mail: howardqi@mtu.edu

S. Liu
Washington State University-Vancouver, Vancouver, WA, USA
e-mail: liusx@vancouver.wsu.edu

C. Wu
University of Missouri-Columbia, Columbia, MO, USA
e-mail: wuchu@missouri.edu

1 Introduction

It is generally believed that most structural models under predict spreads of investment-grade bonds and over predict spreads of junk bonds substantially. However, a careful survey of the literature reveals that the performance of these models is mixed. Among them,[1] the Leland–Toft model (hereafter the LT model) is perhaps the most controversial. Some studies show that it over predicts spreads, in some cases the magnitude of over prediction can be over 5,000 bps (e.g., see Eom et al. 2003). Other studies (e.g., Huang and Huang 2003) show it generates extremely low spreads for investment-grade bonds, and the model performs just as poorly as others.

One reason for these conflicting results is that the LT model has a more complex structure than other models. Another, perhaps more important, reason is the differences in the procedure of calibrating the model. Different calibration procedures can produce substantially different results. Thus, structural models must be properly calibrated to generate dependable results.

In this paper, we propose a rigorous calibration procedure based on a sound theoretical foundation. We then apply the properly calibrated LT model across rating categories and to individual bonds. For comparison purposes, we also apply the Collin-Dufresne and Goldstein (2001) model (hereafter the CG model) to yield data across rating categories. Throughout this paper, *term structure* of spreads refers to yield differences between corporate and Treasury bonds across maturities.

Our paper is related to several studies in the literature of structural models. Leland (1994) proposes a structural pricing model with endogenously determined bankruptcy trigger for risky perpetual bonds. Later, Leland and Toft (1996) extend this model to finite maturity bonds. The LT model considers the firm's capital structure in an environment where there are corporate tax benefits and bankruptcy costs. Within this framework, equity holders optimize the capital structure by finding a bankruptcy trigger such that the equity and firm value are maximized. In essence, the LT model belongs to the traditional family of trade-off theory.

Among the popular models such as the CG model and Longstaff and Schwartz (1995) model (hereafter the LS model), the LT model has perhaps the strongest structural component. Both the CG and LS model assume an exogenously given leverage ratio. Thus, these two models pose no constraint on the leverage ratio. By contrast, the structure of the LT model links default probability to firm's optimal capital structure characteristics, which are determined endogenously by the model. Leverage is thus no longer arbitrarily given; instead it is optimized by the model. The default probability, which is used to calculate the credit spread of a bond, may be regarded as a byproduct of such optimization. We examine the impacts of these different model structures and the calibration procedure on yield spread predictions.

This paper contributes to the literature in three aspects. First, we clarify the theoretical differences between calibrating the LT model and other structural models with exogenous leverage. Second, our paper resolves some of the controversies regarding the explanatory power of the LT model and sheds new light on the structural model

[1] For example, the Collin-Dufresne and Goldstein model (2001), the Longstaff and Schwartz (1995) model, etc.

performance. Third, we apply the calibration procedure to both rating categories and individual bonds to provide additional information regarding the LT model's performance. Our results show that the performance of the LT model can be significantly improved if calibration is done properly.

The remainder of this paper is organized as follows. Section 2 explains the proper procedure of calibration. In particular, we point out a few caveats that one must heed in the model calibration, which are relevant to calibrating any structural models. Section 3 reports the empirical results for both CG and LT models. Discussions and analyses of these results then follow. Finally, Sect. 4 concludes the paper. The technical details of the LT and CG models are summarized in Appendices A and B.

2 Model and calibration

2.1 Why calibration is necessary?

Although most structural models have been criticized for their inability to explain yield spreads in a satisfactory manner (see, for example, Elton et al. 2001; Huang and Huang 2003; Jones et al. 1984), there is little consensus on the performance of these models in the literature. For example, Huang and Huang (2003) and Collin-Dufresne et al. (2001) find that structural models explain only a small portion of the yield spread. On the other hand, Eom et al. (2003) show that the performance of these term structure models is quite mixed.[2] Using the Leland model for perpetual bonds as an approximation for 10-year bonds, Huang and Huang (2003) find that the predicted spreads are fairly small, around 8.9 and 12.4 bps for Aaa and Aa-rated 10-year bonds, respectively. Based on this finding, they conclude that if the finite-maturity version is used, the LT model would perform as poorly as other structural models. In contrast, Eom et al. (2003) find that the LT model is an exception to most structural models in that it over predicts spreads on most bonds.[3] A potential source of these conflicting findings about the performance of the LT model is the calibration procedure. To evaluate the model performance, one needs to estimate how much of the observed yield spread is explained by the model by first assuring that the model-implied default probability is reasonably close to the observed value for bonds with similar credit risk. A proper calibration is necessary to match the model-implied default probability with the observed data to produce reasonable results and to compare different models consistently.

2.2 The calibration process

In a structural model, key financial variables are linked in a certain fashion. The essence of calibration is to tune certain unobserved variables such that the model generates

[2] They report that "most structural models predict spreads that are too high on average except for the Merton model. In particular, they tend to severely overstate the credit risk of the firm with high leverage or volatility and yet suffer from a spread underprediction problem with safer bonds".

[3] In their study, the LT model is shown to predict a spread as high as 5,096 bps while the actual spread is 465 bps.

Table 1 Target parameters for model calibration

Credit rating	Target parameters			Average yield spread (bps)		
	Leverage ratio (%)	Equity premium (%)	Cumulative default probability (%)			
Panel A: 10-year time to maturity						
AAA	13.1	5.38	0.73	63		
AA	21.2	5.60	0.91	91		
A	32.0	5.99	1.96	123		
BBB	43.3	6.55	4.96	194		
BB	53.5	7.30	19.48	320		
B	65.7	8.76	39.96	470		
Panel B: Historical cumulative default rates (%), 1970–1993						
Years	Aaa	Aa	A	Baa	Ba	B
3	0.00	0.08	0.28	0.91	6.92	20.38
5	0.12	0.32	0.62	1.97	11.85	28.38
7	0.33	0.52	1.06	3.09	15.33	34.32
10	0.73	0.91	1.96	4.96	19.48	39.96
20	2.63	2.48	5.23	11.70	29.76	45.58

This table lists the target parameters for model calibration. The leverage ratio parameter is not used for the LT model calibration. The fifth column shows the observed average yield spread. The default probability data are from Fons (1994) and the rest are from Huang and Huang (2003) for the period of 1973–1993. The original data sources are Lehman bond index and Moody's

For the LT model, we set the recovered amount to $\min[0.8\,V_B,\ 0.45P]$, i.e., the lower of 80% of the firm value at bankruptcy and 45% of the face value, a treatment similar to (Huang and Huang (2003)) and Eom et al. (2003). According to Andrade and Kaplan (1998), the cost of financial distress β is in the range of 15–20% of the firm value at bankruptcy. On the other hand, based on the face value measurement, the recovery rate is in the neighborhood of 40–50% (see Federal Reserve publications). Due to the lower bound of 45% face value, the implied bankruptcy cost ratio β ranges from 0.3 to 0.5

Source: Fons (1994). Default probability in panel B is used to implement the LT model for individual bonds

the remaining variables to match the historical data (e.g., default rate). A typical set of calibration targets are listed in Table 1. These target variables are similar to Huang and Huang (2003). However, we note that some variables generated by the model can never match the observed ones. For example, it is well known that trade-off models on capital structure cannot accommodate the observed average leverage ratios across ratings.[4] Therefore, to calibrate a model like the LT model, one must be aware that leverage is not a variable to "calibrate" to match the observed value. The existence of a wide distribution of leverage ratios in each rating category[5] implies that (1) using a

[4] This is supported by a vast volume of literature, including both theoretical and empirical studies, as well as actual industry surveys. For example, see Hittle et al. (1992), Myers (1993) and Pinegar and Wilbricht (1989). Perhaps the most famous argument is the horse-and-rabbit stew analogy by Miller (1977) which points out the much lowered leverage than predicted by the trade-off between corporate tax shields and bankruptcy cost. Myers (1984) proposes the pecking order theory which predicts no particular optimal capital structure at all. However, an in-depth discussion of capital structure is beyond the scope of this paper.

[5] This can also be seen from the statistics of our sample given in Table 2.

Table 2 Sample statistics

	Min	Average	Median	Max		
Panel A: Firm value and leverage distribution of the sample of 286 bonds						
Firm value (million dollars)	98	8,927	2,993	203,416		
Leverage (%)	36	67	66	93		
Panel B: Number of bonds in each category						
Time to maturity	3 years	5 years	7 years	10 years	20 years	Total
AAA-rated (leverage range)					**1**(53%)	1
AA-rated (leverage range)	**3**(63–66%)	**2**(57–66%)		**5**(53–64%)		10
A-rated (leverage range)	**7**(63–72%)	**22**(45–82%)	**33**(53–93%)	**31**(37–76%)	**7**(53–79%)	110
BAA-rated (leverage range)	**17**(39–87%)	**28**(39–87%)	**44**(36–87%)	**50**(39–93%)	**1**(75%)	140
Junk bonds (leverage range)	**5**(62–75%)	**9**(39–83%)	**11**(45–83%)	**10**(62–88%)		35
Total	32	61	88	96	9	286

This table reports the sample statistics. The 286 senior bonds issued during the period of 1980–1992 are chosen from Warga database. Bond credit ratings are based on Moody's and S&P's. The corresponding information on equity return, earnings, payout and leverage is obtained from CRSP/COMPUSTAT. Riskfree rates matching time to maturity and observation times are collected from the Federal Reserve website

Note: The percentages in the parentheses indicate the minimum and maximum of the leverage ratio for the bonds in the corresponding categories. The bond rating and leverage ratio have very poor correlation for the individual bond sample. This is another reason why leverage should not be chosen as a target for calibrating the LT model using rating-based information because the average leverage of a rating category has little meaning if the dispersion is so large

single rating-based average leverage as a target is unrealistic; and (2) there exist many firms operating under sub-optimal capital structures for various reasons.

Although a recent study of corporate debt structure by Hackbarth et al. (2007) suggests that the trade-off theory is sufficient to explain many stylized facts, a survey of the Fortune 500 firms by Pinegar and Wilbricht (1989) finds only 31 percent use target (i.e., optimal) capital structure. Hittle et al. (1992) surveyed 500 large OTC firms and found only 11 percent of these firms target an optimal capital structure. This naturally raises a question—does it still make sense to calibrate the LT model using other rating-based information? Our answer to this question is affirmative.

It is important to note that what ultimately determines the rating boils down to the default probability and recovery rate when bankruptcy occurs. Leverage can only have an effect on bond prices through its effect on default probability. Suppose bonds A and B have the same recovery rate and default probability. Firm A strictly maintains an optimal leverage but firm B does not. Since investors would get the same expected cash flows per dollar invested from both bonds, the two bonds would be priced equally and classified in the same rating category. Therefore, if the LT model can explain the spread of bond A after it is calibrated to match its default probability, it should also be able to explain the spread of bond B. The only difference is that the LT model can explain firm A's leverage but not firm B's.

This example illustrates the importance of choosing appropriate targets for a proper calibration. To calibrate the LT model, we only need to match the observed default probability and recovery ratio. By contrast, in dealing with the CG and LS models, a third dimension, i.e., leverage, must be included in the calibration. Our calibration procedure thus involves two key steps. First, we select a few target parameters to which our model must conform. These parameters must be observable. Second, because we are often facing more target parameter choices than the model can accommodate,[6] we need to identify which are the most important parameters to choose.[7] A proper calibration must ensure that the *links* between the variables are consistent with the model structure.

2.3 Targets for the calibration

In this section, we explain the procedure for model implementation. Based on the structure of the LT model, we choose three targets for calibration: default probability, equity risk premium, and recovery rate.[8] The leverage ratio is added as another target for the CG model's calibration. Table 1 lists these calibration targets. For the CG model, the recovered amount at bankruptcy is assumed to be 45% of the face value. The choice of this recovery rate is in line with the literature (see, for example, Collin-Dufresne and Goldstein 2001; Huang and Huang 2003; Eom et al. 2003, and the government report).[9] In the CG case, the recovery rate is a trivial target since it is directly input into the model. For the LT model, we choose the recovered amount to be *the lower of* 45% of the face value and 80% of the firm value at bankruptcy, V_B (see Andrade and Kaplan 1998). This is because the bankruptcy loss is not measured directly as a fraction of the face value of the debt in the LT model. The model is iterated to determine the correct recovery rate measured by the percentage of the firm value at bankruptcy.

Theoretically, unobserved asset volatility and observable equity volatility are linked through leverage. Shiller (1981) shows that firms' equity return volatility is too high to be justified by the fluctuation in the firm's asset values. Therefore, we choose to adjust asset volatility so that the model-inferred variables match the targets.

2.4 Calibration procedure

The first step for the LT model is to input target recovery ratio and to choose an initial estimate of asset volatility σ. Based on these parameters, the model generates the optimal leverage l_o as well as other variables including yield spread, the values of debt D, equity E, coupon C, bankruptcy threshold V_B, principal P, and the firm value W.

[6] That is, the number of constraints (i.e., targets) cannot exceed the number of unknowns (parameters to be calibrated).

[7] For example, trade-off type models are not suitable to explain observed ratings–leverage relationship because most large profitable firms exhibit strong pecking-order-theory behavior, while some classes of smaller firms show trade-off-theory behavior. Thus if we use data classified by ratings, then the average leverage is not a proper target for our model because there will be too much "noise" from firms who follow different capital structure policies. However, rating-related default probability is a good target because firms with the same rating will most likely have similar default probabilities. Same may be said of recovery rate.

[8] Actually, if we added one more target, the structure of the model would be broken, i.e., the link among some variables specified by the model would not be maintained in all cases. In other words, it is "over-specified".

[9] See http://www.federalreserve.gov/pubs/feds/2005/200510/200510pap.pdf.

The first step needs to be repeated several times with a different recovery ratio to ensure that the recovered amount is the lower of $0.8V_B$ and $0.45P$. For example, if $0.8V_B > 0.45P$, we would choose a lower recovery ratio $(1-\beta)$ until $(1-\beta)V_B = 0.45P$ is satisfied, where β is the bankruptcy cost ratio. The second step is to use the model-inferred optimal leverage l_o and target equity risk premium to calculate the asset risk premium. Based on the MM's model for equity return r_E,[10] we derive the following approximation formula[11] to link equity risk premium $r_E - r_f$ to asset risk premium $r_A - r_f$:

$$r_A - r_f \approx \frac{\left(r_E - r_f\right) + (1-\tau_C) \times \frac{l_o(r_D - r_f)}{1-l_o}}{1 + (1-\tau_C) \times \frac{l_o}{1-l_o}}, \tag{1}$$

where r_A is the expected asset return or the opportunity cost of capital, r_f is the riskfree rate, τ_C is the corporate tax rate, and r_D is the cost of debt capital, which equals c/p for a par bond. The third step is to use the calculated asset premium, the initial asset volatility estimate σ (from the first step) and the relationship $F(t) = N[h_1(t)] + \left(\frac{V_B}{V}\right)^{2a} N[h_2(t)]$ to determine the cumulative default probability under the physical probability measure. The fourth step is to check whether this model-inferred true default probability agrees with the historical default rate. If not, we restart with a different trial asset volatility σ and repeat this process until the two default probabilities, model-implied and observed, agree. In the end, all target parameters are matched and *the structure of the model is perfectly preserved*. With this calibration procedure, there is only one parameter to calibrate at each step, thus preventing the situation that two model-inferred parameters have to be adjusted to match the observed data. This avoids the potential problem of indetermination.

For the CG model, leverage is exogenously given. In the first step, the leverage ratio, recovery rate and equity risk premium are directly input into the model and so matching these targets is trivial. In the second step, we choose the asset volatility σ such that the model would generate the same default probability as the observed value (with the same time horizon). In the third step, we use the same σ with the equity (hence asset) risk premium removed from the model to regenerate default probability under the risk neutral measure. Finally, we use the risk neutral probability to valuate the bond (see (B5) in the Appendix). Since the model uses riskfree zero coupon bond $D^T(r_0)$ as the numeraire, the probability measure is also called the forward measure.

3 Empirical implementation of the model

Our empirical implementation consists of two parts. We first apply the LT and CG model to the same rating categories of 10-year bonds. We then supplement the analysis by applying the LT model to individual bonds with different maturities.

---

[10] For example, see Brealey and Myers (1996), p. 536.

[11] The approximation comes from treating the tax shield as independent of default risk.

3.1 Data description

The 10-year bond data across rating categories are given in panel A of Table 1, where spreads on investment grade bonds are based on the Lehman bond index from 1973 to 1993. Historical junk bond yield spreads were taken from the estimates by Caouette et al. (1998). The default frequencies are from Fons (1994) over the same period. The equity premium was based on regression estimates of Bhandari (1988).

In panel B of Table 1, we list rating-based default rates (from Fons 1994) for five time horizons: 3, 5, 7, 10 and 20 years. These default rates are used to calibrate the LT model for individual bonds in the second part of our empirical analysis. Since for an outstanding bond, there is no "observed" default rate, we use the rating-based default rate as a benchmark for model-generated default probabilities. As we indicate in Sect. 2.2, bonds in the same rating category may have very different leverages, but they should have similar default probabilities.

We also collect a sample of 286 individual senior bonds from the Warga database issued during the period of 1980–1992 in an attempt to match the time period of Fons (1994) default rate data, 1970–1993. Only senior debts are chosen to avoid unnecessary complications associated with junior debts. The corresponding information on equity returns, earnings, payouts and leverage is obtained from CRSP/COMPUSTAT. Riskfree rates matching maturities and observation time are collected from Federal Reserve website. Table 2 summarizes the relevant features of these bonds. Our sample covers a wide range of firm size ($98 million–$203,416 million) and leverage ratios (36–93%).

3.2 Empirical results on 10-year bonds across rating categories

Table 3 reports the credit spreads predicted by the model after calibration. Panel A is for the LT model, while panel B is for the CG model. We report the model-predicted spreads in basis points and percentages of the observed average spreads (in parentheses).

We choose different interest rates and payouts. As demonstrated in this sensitivity analysis, the calibrated models are insensitive to these choices. The fact that calibration produces fairly robust results suggests that the interest and payout choices are not that critical. Our choice of interest rate $r = 4.5$–9% and payout $\delta = 4$–7% covers the values used by Huang and Huang (2003) who choose $r = 8\%$ and $\delta = 6\%$ for the same sample period.

Our results show that a properly calibrated LT model (for finite maturities) has much better performance than reported in previous studies. For example, the estimated credit spread is in the range of 26–30 bps for the AAA bonds. For AA bonds, it increases to a range of 30–37 bps. They are significantly higher than 8.9 and 12.4 bps reported in Huang and Huang (2003) for AAA and AA bonds, respectively. Note that Huang and Huang (2003) use the perpetual Leland (1994) model to approximate the Leland and Toft (1996) model for finite maturities. Their results with the perpetual model appear to be better. However, they note that once the true LT model is used, the performance will drop considerably. Another significant improvement is that the calibrated model

Table 3 Model predicted spreads versus observed average spreads

Panel A: The LT model with 10-year time to maturity

	Recovered amount is min $[0.8\ V_B,\ 0.45P]$*					
	AAA	AA	A	BBB	BB	B
Average yield spread (bps)	63	91	123	194	320	470
$r = 4.5\%$; $\delta = 4.0\%$	26 (41%)	30 (33%)	46 (37%)	84 (43%)	229 (72%)	374 (80%)
$r = 6.0\%$; $\delta = 4.0\%$	29 (46%)	35 (38%)	60 (49%)	96 (49%)	237 (74%)	374 (80%)
$r = 6.0\%$; $\delta = 5.5\%$	32 (51%)	35 (38%)	53 (43%)	92 (47%)	235 (73%)	371 (79%)
$r = 7.5\%$; $\delta = 7.0\%$	35 (56%)	39 (43%)	57 (46%)	97 (50%)	242 (76%)	380 (81%)
$r = 9.0\%$; $\delta = 7.0\%$	30 (48%)	37 (41%)	60 (49%)	101 (52%)	248 (78%)	392 (83%)

Panel B: The CG model with 10-year time to maturity

	Recovered amount is $0.45P$					
	AAA	AA	A	BBB	BB	B
Average yield spread (bps)	63	91	123	194	320	470
$r = 4.5\%$; $\delta = 4.0\%$	4 (6%)	6 (7%)	12 (10%)	32 (16%)	126 (39%)	290 (62%)
$r = 6.0\%$; $\delta = 4.0\%$	4 (6%)	6 (7%)	12 (10%)	32 (16%)	130 (41%)	297 (63%)
$r = 6.0\%$; $\delta = 5.5\%$	4 (6%)	6 (7%)	12 (10%)	33 (17%)	131 (41%)	298 (63%)
$r = 7.5\%$; $\delta = 7.0\%$	4 (6%)	6 (7%)	12 (10%)	32 (16%)	136 (43%)	301 (64%)
$r = 9.0\%$; $\delta = 7.0\%$	4 (6%)	6 (7%)	12 (10%)	32 (16%)	135 (42%)	308 (66%)

This table reports the credit spreads predicted by the LT and CG models after calibration. Cases with different interest rates and payouts are tested. The corporate tax rate τ_C is 35%. We report the spreads in basis points and as the percentage of the observed average spread in parentheses

* The recovered amount is chosen to be the lower value of 80% of the firm value at bankruptcy and 45% of the face value of debt

does not generate excessively high spreads, as documented by Eom et al. (2003). However, we note that our results and theirs may not be compared directly since the sample periods and bond screening criteria are different. Specifically, we only select senior debts, while their sample is screened through different and more criteria.

Compared to the CG model, the calibrated LT model has considerably greater explanatory power across all rating categories. For example, the CG model explains only 6% of the observed spreads for AAA bonds, and 16% for BBB bonds, while the LT model explains about 50% of the observed spread. Even though the CG model's explanatory power increases significantly for junk bonds (e.g., from 6 to 16% for investment grade bonds to 39–66% for junk bonds), the calibrated LT model still has considerably better performance, explaining 72–83% of the observed spreads for junk bonds.

Our findings raise a question of whether the LT model performs better than the CG model.[12] Although our results seem to suggest this way, we could only claim

[12] The current general belief is that all the structural models have roughly equally poor performances after calibration across rating categories. We show that if the true LT model (for finite maturities) is properly calibrated, the situation may be different from previously believed.

that compared with the CG model, the LT model has significantly greater explanatory power for rating-based categories (or portfolios). This is because within each rating category, firms can have very different leverages (see Table 2, panel B), which makes a single average leverage much less meaningful as a calibration target. The CG model needs leverage as a calibration target, while the LT model's calibration does not use rating-based leverage. Therefore, the LT model is expected to be free from this source of error and becomes more reliable (see Sect. 2.2 for why the LT model may be calibrated using the rating-based information). The CG model does not suit the portfolios based on rating categories well. We expect the CG model to have a greater explanatory power if bond portfolios are formed by similar leverages, instead of rating groups. We leave this subject for future investigation.

3.3 Applying the LT model to individual bonds: empirical results and analysis

Table 4 reports the credit spreads predicted by the calibrated LT model for individual bonds. In panel A, we report the model explanatory power (i.e., the model-predicted spread as a percentage of the actual spread) for four portfolios constructed by firm size. The number of bonds in each portfolio is about the same. For example, column (2) includes the portfolio consisting of 71 bonds with the largest market capitalization in our sample of 286 bonds. We calibrate the LT model for every individual bond and report the asset-weighted average explanatory power for each portfolio.

Our results in panel A suggest that firm size has little effect on the LT model's performance. The difference is very small among all asset-based portfolios. This is most likely because the distribution of credit ratings is similar within each portfolio, suggesting that bond quality does not depend on firm size.

In panel B of Table 4, we construct two leverage-based portfolios with an equal number of bonds. One contains the 143 bonds with low leverage in column (2) and the other with high leverage in column (3). Our results show a difference. The LT model has an average explanatory power of 27.4% for the lower levered portfolio, and 36.9% for the higher levered portfolio. However, this difference may be explained by the fact that the two portfolios are constructed to implicitly approximate the rating-based categories (e.g., see Table 3). The higher levered portfolio contains lower-rated bonds. Since the LT model has a greater explanatory power for junk bonds than for investment-grade bonds (see Table 3), it is not surprising that it performs better for the higher-leverage portfolio.

In panel C, we further separate the bonds into portfolios based on their rating and time to maturity. Since the number of junk bonds is too small for some time horizons, we focus on investment grade bonds. First of all, we do not observe a general trend in the relationship between model explanatory power and time to maturity. This is perhaps due to the small sample size of the portfolios or perhaps there is indeed no such simple relationship. However, there is an apparent difference between the portfolios with a rating higher than A and the BBB-rated portfolios. In our sample, the former contains predominantly A and AA bonds. As shown in panel A of Table 3, the LT model explains BBB bonds better than AA and A bonds, although the difference is not as big as shown in panel C of Table 4.

Table 4 The explanatory power of the LT model for individual bonds

(1)	(2)	(3)	(4)	(5)	(6)
Panel A[a]					
Asset-based portfolios (number of bonds)	0–25% quantile (71)	25–50% quantile (72)	50–75% quantile (71)	75–100% quantile (72)	Entire sample (286)
Asset-weighted average explanatory power	36.86%	32.80%	32.89%	35.19%	34.83%
Panel B[b]					
(1)	(2)	(3)	(4)		
Leverage-based portfolios	$l = 36$–67% (143 bonds)	$l = 67$–93% (143 bonds)	Entire sample (286 bonds)		
Asset-weighted average explanatory power	27.40%	36.87%	34.83%		
Panel C[c]					
Panel C1. T = 3 years					
(1)	(2)	(3)	(4)		
Bond rating	A-rated or higher (10 bonds)	BBB-rated (17 bonds)	Investment grade (27 bonds)		
Asset-weighted average explanatory power	10.48%	46.30%	28.49%		
Panel C2. T = 5 years					
(1)	(2)	(3)	(4)		
Bond rating	A-rated or higher (24 bonds)	BBB-rated (28 bonds)	Investment grade (52 bonds)		
Asset-weighted average explanatory power	34.44%	50.75%	38.40%		
Panel C3. T = 7 years					
(1)	(2)	(3)	(4)		
Bond rating	A-rated or higher (33 bonds)	BBB-rated (44 bonds)	Investment grade (77 bonds)		
Asset-weighted average explanatory power	18.02%	34.37%	20.74%		
Panel C4. T = 10 years					
(1)	(2)	(3)	(4)		
Bond rating	A-rated or higher (36 bonds)	BBB-rated (50 bonds)	Investment grade (86 bonds)		
Asset-weighted average explanatory power	19.31%	61.14%	45.28%		

This table reports the explanatory power of the LT model in terms of the model-predicted spreads as a percentage of the actual spreads

[a] *Firm size effect* The total sample of 286 bonds are ranked according to the firm size and separated into four portfolios of roughly same number of bonds (71–72 bonds). Column (2) is for the bond portfolio with largest firm size. Column (5) is for the portfolio with smallest firm size. Column (6) is for the entire sample

[b] *Leverage effect* The total sample of 286 bonds are ranked according to the leverage ratio and separated into two portfolios. Column (2) is for the bond portfolio with lower leverage ratio (36–67%) and column (3) is with higher leverage ratio (67–93%)

[c] *Time to maturity and investment-grade bonds* Junk bonds are not considered since its sample size is too small for certain horizons

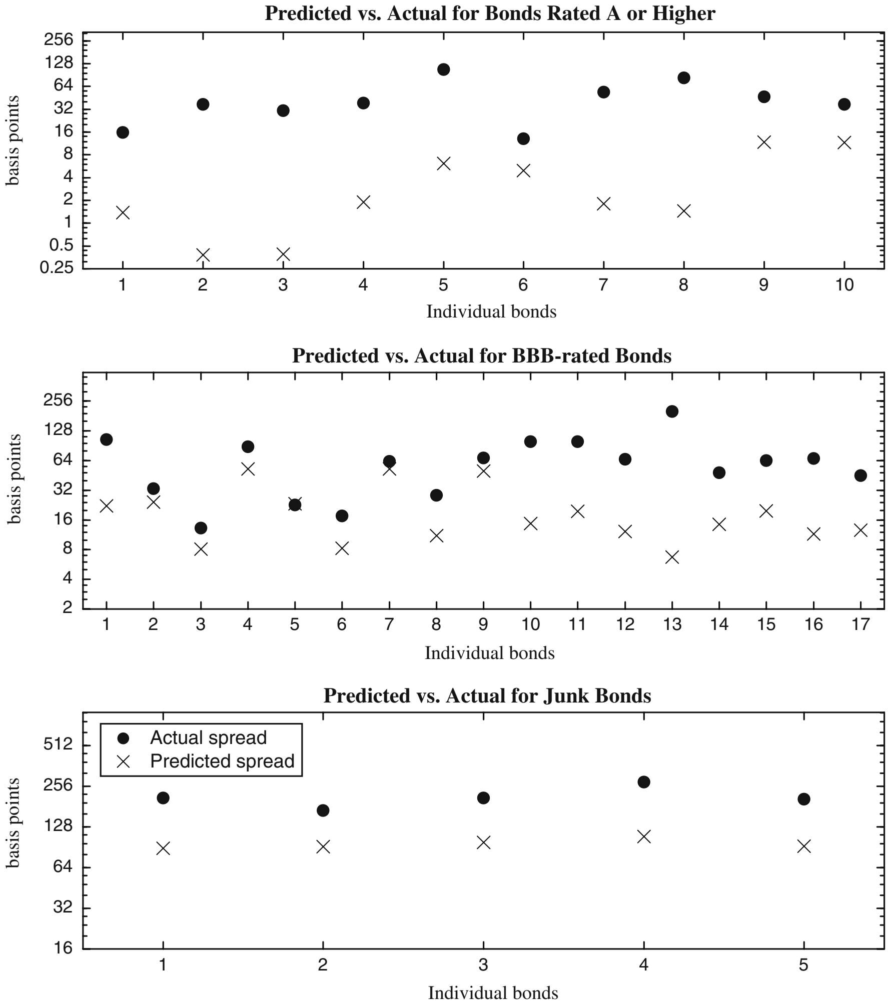

Fig. 1 LT model and individual bonds with maturity of 3 years. This figure plots the predicted and actual spreads for individual bonds. There are 10 bonds rated A or higher (*upper panel*), 17 BBB-rated bonds (*middle panel*) and 5 junk bonds (*bottom panel*)

Figures 1, 2, 3, 4 and 5 plot out our results for each bond in our sample separated into rating- and maturity-based groups. A close examination of these figures and Tables 3 (panel A) and 4 (panel C) suggests that in general the LT model (and the CG model) has a greater explanatory power for lower-rated bonds, especially for junk bonds. This is largely because credit risk plays the most important role in these two models. Other factors, such as personal taxes, are expected to exert a relative stable influence across rating categories since the tax premium is mostly associated with the state taxes.[13]

[13] For non-tax exempt investors, coupons from corporate bonds are subject both state and federal taxes, while coupons from government debts are subject to federal taxes.

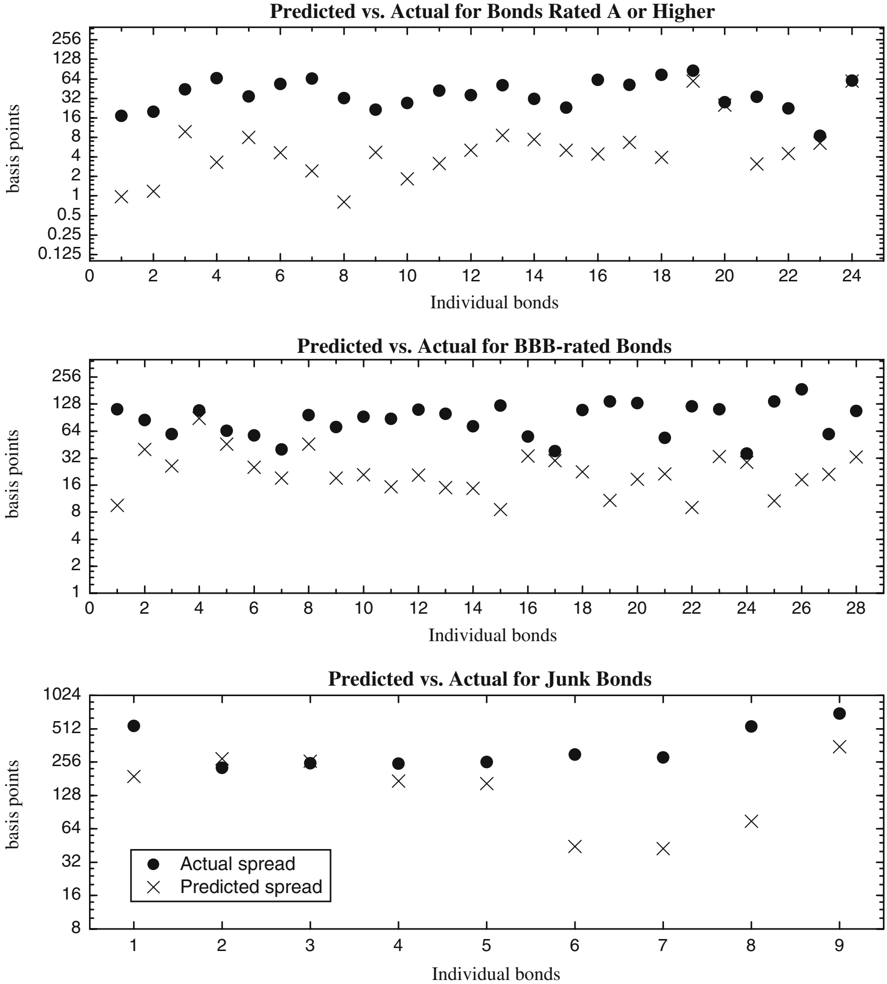

Fig. 2 LT model and individual bonds with maturity of 5 years. This figure plots the predicted and actual spreads for individual bonds. There are 24 bonds rated A or higher (*upper panel*), 28 BBB-rated bonds (*middle panel*) and 9 junk bonds (*bottom panel*)

When rating declines, one would expect credit risk to play a relatively more important role than other most stable factors (such as personal taxes). Hence, the credit risk models would appear to have greater explanatory powers.

4 Conclusions

In this paper, we focus on the proper calibration procedure for different structural models, specifically the LT model. A proper calibration must ensure that all internal *links* inherent in the model are preserved throughout the calibration. When there are

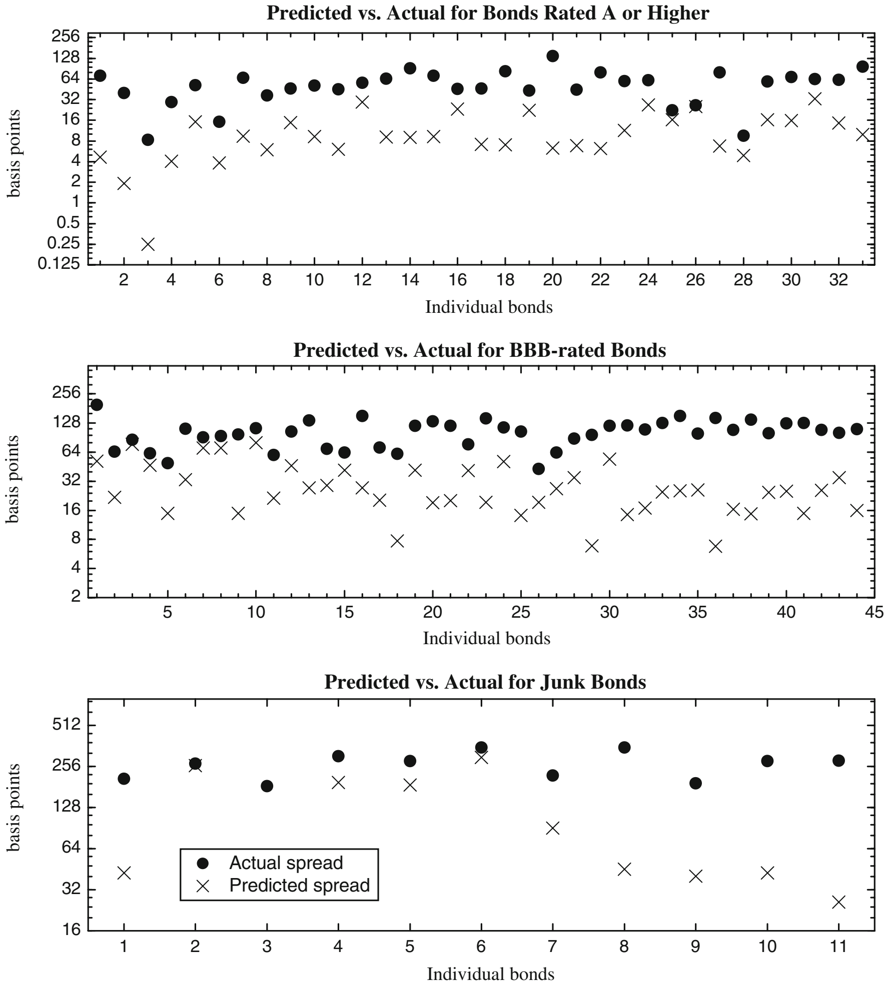

Fig. 3 LT model and individual bonds with maturity of 7 years. This figure plots the predicted and actual spreads for individual bonds. There are 33 bonds rated A or higher (*upper panel*), 44 BBB-rated bonds (*middle panel*) and 11 junk bonds (*bottom panel*)

more target candidates than the model can take, the choice of targets must be in line with the structure of the original model. Otherwise, the problems of spurious calibration and inconsistency with the structure of the model may occur. In particular, the leverage ratio would be a poor target for calibrating the LT model, while it is quite relevant for calibrating the CG model.

Using the calibrated models, we find that the LT model performs significantly better than what has been reported. It has significantly greater explanatory power than the CG model when applied to various rating groups. This is in contrast with previous findings for the performance of structural models. However, this performance difference is

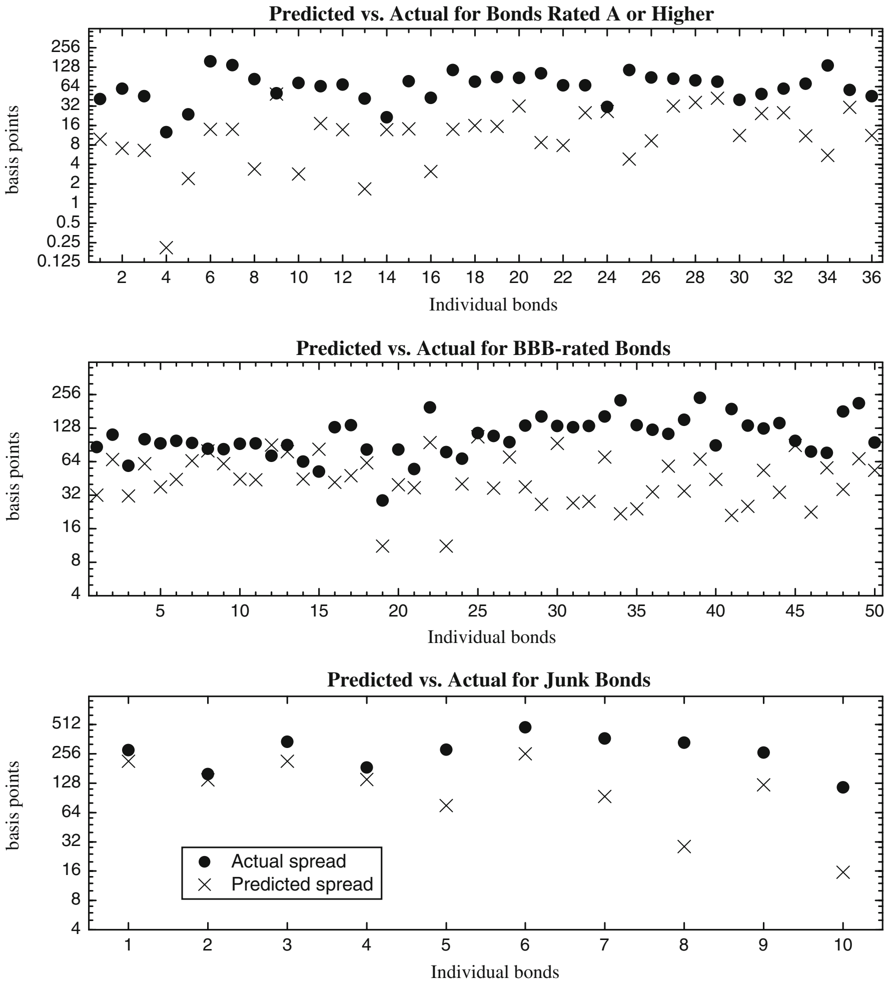

Fig. 4 LT model and individual bonds with maturity of 10 years. This figure plots the predicted and actual spreads for individual bonds. There are ten 36 bonds rated A or higher (*upper panel*), 50 BBB-rated bonds (*middle panel*) and junk bonds (*bottom panel*)

largely due to fact that the LT model is better suited for rating-based portfolios than the CG model because the latter requires the input of an exogenous leverage ratio in the calibration. A single rating-based average leverage is not a suitable target given the wide dispersion of leverage for firms in each rating category. The LT model is free from this problem, since it avoids using an exogenous leverage ratio as a target in its calibration. We also extend our empirical analysis by applying the calibrated LT model to 286 individual senior bonds where portfolios are formed based on ratings, firm size and maturities. The results from the individual bond sample are consistent with our general findings and provide additional insights. Our results suggest that the LT model performs quite well when it is calibrated properly.

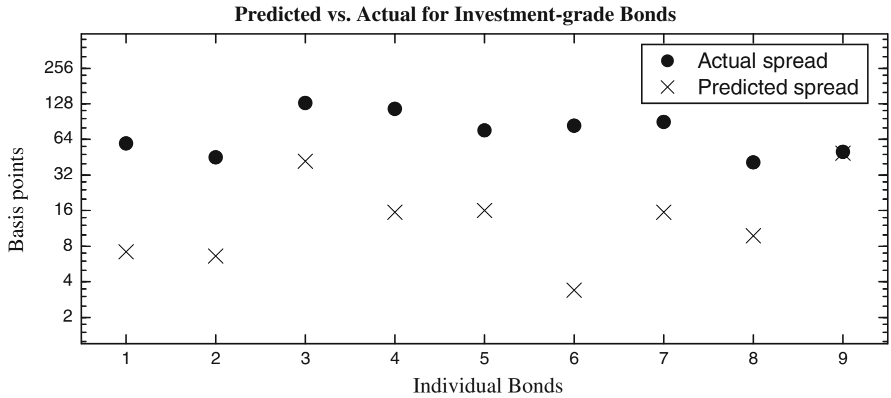

Fig. 5 LT model and individual bonds with maturity of 20 years. This figure plots the predicted and actual spreads for individual bonds. There are nine investment-grade bonds

Acknowledgments We thank Dr. Janet Payne, the discussant at the 2004 FMA Annual Meeting, and the seminar participants at Michigan Tech and Syracuse University for helpful discussions. We benefit intellectually from the insightful and thorough comments by an anonymous referee and the editor, which significantly improved this study.

Appendix A

Brief review of the Leland–Toft model

A.1. General setting

By convention,[14] the asset of an unlevered firm, V, is postulated to have the following diffusion process:

$$\frac{dV}{V} = [\mu(V, t) - \delta]\, dt + \sigma dZ, \tag{A.1}$$

where $\mu(V, t)$ is the total expected rate of return on value V; δ is the payout rate; σ is the volatility and is a constant. Bankruptcy happens when the firm value hits a boundary, V_B, for the first time. Assume the firm continuously issues new debt d and redeems debt that expires. Within unit time, coupon flow is c and principal flow is p. Upon bankruptcy, debt holders receive a fraction ρ of the asset value V_B. By

[14] For example, see Black and Cox (1976), Brennan and Schwartz (1978), Leland and Toft (1996) and Merton (1974).

risk-neutral valuation, the value of the bond is given by

$$d(V, V_B, t) = \int_0^t c(t) e^{-rs} [1 - F(s, V, V_B)] ds + p(t) e^{-rt} [1 - F(t, V, V_B)] + \int_0^t \rho V_B e^{-rs} f(s, V, V_B) ds \tag{A.2}$$

where t is time to maturity, $F(s, V, V_B)$ is the cumulative default probability up to time s, and $f(s, V, V_B)$ is the incremental default probability from time s to $s + \Delta s$. Leland and Toft (1996) gives the following solution

$$d(V, V_B, t) = \frac{c(t)}{r} + e^{-rt} \left[p(t) - \frac{c(t)}{r} \right] (1 - F(t)) + \left[\rho(t) V_B - \frac{c(t)}{r} \right] G(t), \tag{A.3}$$

where $F(t) = N[h_1(t)] + \left(\frac{V_B}{V}\right)^{2a} N[h_2(t)], G(t) = \left(\frac{V_B}{V}\right)^{a-z} N[q_1(t)] + \left(\frac{V_B}{V}\right)^{a+z} N[q_2(t)]$ and $N(\cdot)$ denotes the cumulative standard normal distribution. Other parameters are

$$q_1(t) = \frac{-b - z\sigma^2 t}{\sigma\sqrt{t}}; \quad q_2(t) = \frac{-b + z\sigma^2 t}{\sigma\sqrt{t}};$$
$$h_1(t) = \frac{-b - a\sigma^2 t}{\sigma\sqrt{t}}; \quad h_2(t) = \frac{-b + a\sigma^2 t}{\sigma\sqrt{t}}$$
$$a = \frac{r - \delta - (\sigma^2/2)}{\sigma^2}; \quad b = \ln\left(\frac{V}{V_B}\right); \quad z = \frac{[(a\sigma^2)^2 + 2r\sigma^2]^{1/2}}{\sigma^2}$$

The LT model assumes that the firm continuously issues a constant principal amount of new debt with maturity T and the same amount of debt is retired. Thus, the debt structure is stationary in continuous time. The value of all outstanding bonds D can be determined by integrating the debt flow $d(V, V_B, t)$ over a period of T:

$$D(V, V_B, T) = \int_{t=0}^{T} d(V, V_B, t) dt, \tag{A.4}$$

A.2. The optimal capital structure

Given the asset value $V(t)$, tax benefits $h(t)$, bankruptcy costs $B(V)$, and the total outstanding debt $D(V, V_B, T)$, the levered firm's value $W(V, V_B, T)$ is given by

$$W(V, V_B, T) = V + h(V) - B(V) \tag{A.5}$$

where the tax benefits $h(V) = \frac{\tau_C C}{r}\left[1 - \left(\frac{V_B}{V}\right)^{a+z}\right]$ and $B(V) = \beta V_B \left(\frac{V_B}{V}\right)^{a+z}$ are given in Leland (1994). The equity value $E(V, V_B, T)$ is given by

$$\begin{aligned} E(V, V_B, T) &= W(V, V_B, T) - D(V, V_B, T) \\ &= V + h(V) - B(V) - D(V, V_B, T) \\ &= V + \frac{\tau_C C}{r}\left[1 - \left(\frac{V_B}{V}\right)^{a+z}\right] - \beta V_B \left(\frac{V_B}{V}\right)^{a+z} - D(V, V_B, T) \end{aligned} \tag{A.6}$$

Bankruptcy occurs once the firm can no longer service its debt. The default boundary V_B can be solved by using the smooth-pasting condition

$$\left.\frac{\partial E(V, V_B, T)}{\partial V}\right|_{V=V_B} = 0 \tag{A.7}$$

Without loss of generality, we can set the initial asset value $V = 100$, and impose par-bond constraint $d(V, c, p)|_{V=100} = p(T)$ to eliminate p. This leaves us V_B as a function of c. The smallest coupon density c (i.e., annual coupon payment $c = C/T$) that satisfies par-bond constraint is regarded as the solution to the optimal leverage problem.

Appendix B

The Collin-Dufresne and Goldstein model

The CG model is characterized by the following processes for firm value V_t, stochastic interest rate r_t, and log-default threshold k_t:

$$\frac{dV_t}{V_t} = (r_t - \xi)dt + \sigma dz_1^{\Pi}(t), \tag{B1}$$

$$dr_t = \beta(\theta - r_t)dt + \eta dz_2^{\Pi}(t), \tag{B2}$$

$$dk_t = \lambda[y_t - \nu - \phi(r_t - \theta) - k_t]dt, \tag{B3}$$

The price of a risky tax-free discount bond is given by:

$$P^T(r_0, l_0) = D^T(r_0)\ [1 - w\,Q^T(r_0, l_0, T)] \tag{B4}$$

where w is the fractional loss in principal, $D^T(r_0)$ is the price of the risk-free zero-coupon bond with maturity T and $Q^T(r_0, l_0, T)$ is the cumulative default probability at time T under the T-forward measure which is

$$Q^T(r_0, l_0, T) = \sum_{j=1}^{n_T}\sum_{i=1}^{n_r} q(r_i, l_0, t_j), \tag{B5}$$

where $q(r_i, l_0, t_j)$ is the probability mass in a grid $(\Delta r \times \Delta t)$ at the level of (r_i, t_j), and

$$q(r_i, l_0, t_1) = \Delta r \Psi(r_i, t_1, l_0), \quad \forall i \in (1, 2, \dots, n_r), \tag{B6}$$

$$q(r_i, l_0, t_j) = \Delta r \left[\Psi(r_i, t_j, l_0) - \sum_{v=1}^{j-1} \sum_{u=1}^{n_r} q(r_u, t_v, l_0) \Phi(r_i, t_j | r_u, t_v) \right], \quad \forall i \in (1, 2, \dots, n_r) \;\; \text{and} \;\; \forall j \in (2, \dots, n_T). \tag{B7}$$

The solutions for $\Psi(r_t, t, l_0)$ and $\Phi(r_t, t | r_s, s)$ are given in Collin-Dufresne and Goldstein (2001).

References

Andrade, G., Kaplan, S.: How costly is financial (not economic) distress? Evidence from highly levered transactions that became distressed. J Finance **53**, 1443–1493 (1998)

Bhandari, C.: Debt/equity ratio and expected common stock returns: Empirical evidence. J Finance **43**, 507–528 (1988)

Black, F., Cox, J.: Valuing corporate securities: Some effects of bond indenture provisions. J Finance **31**, 351–367 (1976)

Brealey, R.A., Myers, S.C.: Principles of Corporate Finance, 5th edn. The McGraw-Hill Companies, Inc., New York (1996)

Brennan, M., Schwartz, E.: Corporate income taxes, valuation, and the problem of optimal capital structure. J Bus **51**, 103–114 (1978)

Caouette, J.B., Altman, E.I., Narayanan, P.: Managing Credit Risk: the Next Great Financial Challenge. Wiley, New York (1998)

Collin-Dufresne, P., Goldstein, R.S.: Do credit spreads reflect stationary leverage ratios? J Finance **56**, 2177–2208 (2001)

Collin-Dufresne, P., Goldstein, R.S., Martin, J.S.: The determinants of credit spread changes. J Finance **56**, 2177–2207 (2001)

Elton, E.J., Gruber, M.J., Agrawal, D., Mann, C.: Explaining the rate spread on corporate bonds. J Finance **56**, 247–276 (2001)

Eom, Y.H., Helwege, J., Huang, J.-Z.: Structural models of corporate bond pricing: An empirical analysis. Rev Financ Stud **17**, 499–544 (2003)

Fons, J.: Using default rates to model the term structure of credit risk. Financial Anal J **September–October**, 25–32 (1994)

Hackbarth, D., Hennessy, C.A., Leland, H.E.: Can the trade-off theory explain debt structure? Rev Financ Stud **20**, 1389–1428 (2007)

Hittle, L.C., Haddad, K., Gitman, L.J.: Over-the-counter firms, asymmetric information and financing preferences. Rev Financ Econ **2**(1), 81–92 (1992)

Huang, J.-Z., Huang, M.: How much of the corporate-Treasury yield spread is due to credit risk? working paper, Penn State and Stanford Universities (2003)

Jones, E.P., Mason, S., Rosenfeld, E.: Contingent claims analysis of corporate capital structures: An empirical investigation. J Finance **39**, 611–625 (1984)

Leland, H.: Corporate debt value, bond covenants, and optional capital structure. J Finance **49**, 1213–1252 (1994)

Leland, H., Toft, K.: Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. J Finance **51**, 987–1019 (1996)

Longstaff, F., Schwartz, E.: A simple approach to valuing risky fixed and floating rate debt. J Finance **50**, 789–891 (1995)

Merton, R.: On the pricing of corporate debt: The risk structure of interest rates. J Finance **29**, 449–470 (1974)

Miller, M.: Debt and taxes. J Finance **32**, 261–275 (1977)

Myers, S.C.: The capital structure puzzle. J Finance **39**, 575–592 (1984)

Myers, S.C.: Still searching for optimal capital structure. J Appl Corporate Finance **39**, 4–14 (1993)
Pinegar, J.M., Wilbricht, L.: What managers think of capital structure theory: A survey. Financ Manage **Winter**, 82–91 (1989)
Shiller, R.: Do stock prices move too much to be justified by subsequent changes in dividends? Am Econ Rev **71**, 421–436 (1981)